JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 14 No. 2, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v14i2.9512

MULTIVARIATE TIME-SERIES DATA PREPARATION FOR WATER QUALITY FORECASTING IN LIVING LAB SMART AQUACULTURE SYSTEM

Cindy Hapsari[1*], Ni Putu Novita Puspa Dewi[2], Bagus Gede Krishna Yudistira[3], Gede Defry Widhi Adnyana[4], Putu Zasya Eka Satya Nugraha[5], Komang Ari Widiani[6]

[1,2,3,4]Department of Informatic Engineering, Faculty of Engineering and Vocational, Universitas Pendidikan Ganesha; Jl. Udayana No.11, Banjar Tegal, Singaraja, Kabupaten Buleleng, Bali 81116
[5,6]Dago Engineering;Jl. Bukit Dago Selatan No.27-29, Dago, Kecamatan Coblong, Kota Bandung, Jawa Barat; 40135

Keywords:
Aquaculture, Water Quality Monitoring, IoT, Sensor Data, Time-Series Forecasting, Preprocessing Data, Sliding Window Transformation

Corespondent Email:
ccchhh666888@gmail.com

Abstract. *Water quality monitoring in biofloc aquaculture systems is still commonly performed through manual observation, which limits the ability to detect short-term fluctuations in environmental parameters. Although Internet of Things (IoT)-based sensors enable continuous data acquisition, raw sensor datasets often contain missing values, noise, and inconsistent temporal structures that reduce their suitability for time-series forecasting applications. This study proposes a structured preprocessing pipeline for multivariate water quality sensor data consisting of temperature, pH, and total dissolved solids (TDS) to improve dataset readiness for predictive modeling. The preprocessing stages include data filtering, interpolation, outlier detection, normalization using Min–Max scaling, and sliding window transformation to construct supervised multi-step forecasting sequences with a 144-timestep input–output horizon. Experimental validation using a Long Short-Term Memory (LSTM) model demonstrates that the transformed dataset supports stable forecasting performance across multiple parameters. The proposed preprocessing framework contributes to improving the reliability of IoT-based aquaculture monitoring systems and supports the development of intelligent early warning mechanisms for water quality management.*

Abstract. Pemantauan kualitas air dalam sistem akuakultur biofloc masih umum dilakukan melalui pengamatan manual, yang membatasi kemampuan untuk mendeteksi fluktuasi jangka pendek pada parameter lingkungan. Meskipun sensor berbasis Internet of Things (IoT) memungkinkan akuisisi data berkelanjutan, dataset sensor mentah seringkali mengandung nilai yang hilang, noise, dan struktur temporal yang tidak konsisten yang mengurangi kesesuaiannya untuk aplikasi peramalan deret waktu. Studi ini mengusulkan pipeline pra-pemrosesan terstruktur untuk data sensor kualitas air multivariat yang terdiri dari suhu, pH, dan total padatan terlarut (TDS) untuk meningkatkan kesiapan dataset untuk pemodelan prediktif. Tahapan pra-pemrosesan meliputi penyaringan data, interpolasi, deteksi outlier, normalisasi menggunakan penskalaan Min–Max, dan transformasi jendela geser untuk membangun urutan peramalan multi-langkah terawasi dengan horizon input-output 144 langkah waktu. Validasi eksperimental menggunakan model Long Short-Term Memory (LSTM) menunjukkan bahwa dataset yang ditransformasikan mendukung kinerja peramalan yang stabil di berbagai parameter. Kerangka kerja pra-pemrosesan yang diusulkan berkontribusi pada peningkatan keandalan sistem pemantauan akuakultur berbasis IoT.

1. INTRODUCTION

Smart aquaculture systems integrating Internet of Things (IoT) and Artificial Intelligence (AI) enable real-time monitoring of water quality parameters and support data-driven decision-making to improve productivity and sustainability in fish farming[1]. Water quality plays a critical role in maintaining the stability and productivity of aquaculture systems, particularly in biofloc environments where microbial activity strongly affects fish growth and survival. Continuous monitoring of parameters such as temperature, pH, and total dissolved solids (TDS) is essential to prevent environmental imbalance and production losses in aquaculture operations[2]. However, in many practical aquaculture settings, monitoring activities are still performed manually, which limits the ability to detect short-term fluctuations and reduces the effectiveness of early mitigation strategies for maintaining optimal pond conditions[3].

2. LITERATURE REVIEW

2.1 Internet of Things

Recent developments in Internet of Things (IoT) technology enable continuous and real-time acquisition of environmental parameters through sensor-based monitoring systems. These technologies support data-driven decision making and improve monitoring efficiency in smart aquaculture environments[4]. Nevertheless, raw sensor datasets collected from IoT monitoring platforms frequently contain missing values, noise, and inconsistent temporal structures that reduce their suitability for predictive modeling without proper preprocessing procedures[5].

2.2 LSTM (Long Short Term Memory)

Machine learning and deep learning approaches have been widely applied for forecasting water quality parameters due to their ability to model temporal dependencies in sequential environmental data. In particular, Long Short-Term Memory (LSTM) networks have demonstrated strong performance in multivariate water quality prediction tasks by capturing relationships between multiple environmental variables simultaneously[6]. Several recent studies have also integrated IoT-based monitoring systems with deep learning models to support intelligent aquaculture management and early warning mechanisms for water quality control[7].

2.3 Data Analyst Observation

Despite these advances, many existing studies primarily emphasize prediction accuracy while providing limited discussion on structured preprocessing pipelines required to transform raw multivariate sensor observations into forecasting-ready datasets. Proper preprocessing of time-series sensor data, including handling missing values, noise filtering, normalization, and supervised sequence construction using sliding window techniques, plays an essential role in improving prediction reliability and model stability in environmental monitoring systems[8]. Therefore, this study proposes a structured preprocessing pipeline for multivariate IoT-based water quality sensor data consisting of temperature, pH, and TDS parameters. The proposed approach transforms raw time-series observations into multi-step forecasting sequences using filtering, interpolation, normalization, and sliding window transformation techniques to support predictive modeling and intelligent early warning systems in aquaculture monitoring environments.

2.4 Data Processing Steps

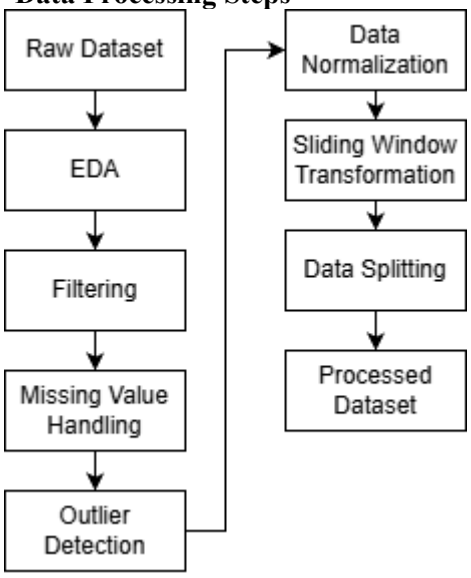

Figure 1. Data Processing Steps

The data processing workflow applied in this study is illustrated in Fig. 1. The process begins with raw dataset acquisition, followed by exploratory data analysis (EDA) to examine data characteristics.

Subsequently, data filtering is performed to ensure temporal consistency, followed by missing value handling to reconstruct incomplete observations. Outlier detection is then applied to remove abnormal values, and data normalization is conducted to standardize the range of each parameter. Finally, the processed data is transformed and divided into training, validation, and testing sets. To support forecasting tasks, a sliding window transformation is applied as the final stage of processing. Sliding window transformation was applied to convert sequential time-series observations into supervised learning sequences suitable for multistep forecasting tasks. This transformation organizes historical observations as input sequences and future observations as prediction targets, enabling forecasting models to learn temporal dependencies from structured multivariate sensor data [9]. The concept of sliding window transformation and dataset framing used in this study is illustrated in Fig. 1, where sequential observations are restructured into input–output pairs for forecasting preparation.

3. RESEARCH METHODOLOGY

This study proposes a structured preprocessing pipeline for multivariate Internet of Things (IoT)-based water quality sensor data to improve dataset readiness for time-series forecasting applications in aquaculture monitoring environments. The preprocessing workflow consists of exploratory data analysis, data filtering, missing value handling, outlier detection, normalization, sliding window transformation, and dataset splitting. These stages transform raw temporal observations into supervised multivariate sequences suitable for multi-step forecasting tasks.

Structured preprocessing plays an important role in improving prediction stability and reliability in environmental time-series forecasting, particularly when dealing with multivariate sensor observations collected over long monitoring periods[10]. Similar preprocessing data strategies have been widely applied in intelligent monitoring systems to ensure consistency and usability of sensor-based datasets before predictive modeling is performed[11] are been used in Fig. 1.

3.1. Data Description

Sensor measurements were recorded at 30-minute intervals over an observation period of approximately 90 days from last September 2025 until early January 2026. The collected data were automatically stored in CSV format and had been calibrated at the sensor level prior to acquisition. Continuous monitoring using IoT-based sensor systems enables efficient collection of temporal environmental data for predictive aquaculture applications[12]. Before sequence transformation was performed (Table 1), the dataset was examined to ensure temporal consistency and completeness across all variables. Multivariate environmental monitoring datasets with similar temporal structures have been widely used in recent forecasting studies to support multi-step prediction of water quality conditions[13].

Table 1. Characteristic of Unfiltered Data

Parameters	Unit	Interval Time
Temperature	°C	15 Secs
pH	-	15 Secs
TDS	ppm	15 Secs

After preprocessing and sequence transformation, a total of 4,174 input–output sequence pairs were generated using multivariate sliding window construction for forecasting preparation. Similar multivariate environmental monitoring datasets have been widely used in recent smart aquaculture prediction studies to support early warning systems and decision-support mechanisms.

3.2. Exploratory Data Analysis (EDA)

Exploratory data analysis (EDA) was conducted to examine the statistical characteristics and inter-variable relationships of the collected water quality dataset prior to preprocessing. This stage is essential for identifying measurement irregularities, detecting abnormal sensor observations, and understanding the distribution patterns of environmental parameters before transforming the dataset into supervised forecasting sequences[14].

Descriptive statistical analysis was performed to summarize the distribution of temperature, pH, and total dissolved solids (TDS) observations across the monitoring period. The raw dataset consisted of 196,871 temperature records, 196,871 pH records, and 196,867 TDS records. The statistical summary indicates that several parameters contain extreme minimum and maximum values outside normal operational ranges of aquaculture environments, such as temperature values reaching 305.625 °C, pH values reaching 757.656, and TDS values exceeding 4,357,650 ppm. The presence of these extreme values in Table 2 suggests measurement inconsistencies and potential noise in the sensor observations, which may affect forecasting reliability if not addressed during preprocessing.

Table 2. Descriptive Statistics of Raw Water Quality Dataset

Statistics	Temperature	pH	TDS
Count	196,871	196,871	196,867
Mean	28.869533	7.485431	298.5504
Std	1.679138	4.519726	22,571.92
Min	0.00	-24.70	-1,288,860
25%	27.75	7.22	51.00
50% (Median)	28.88	7.36	364.00
75%	30.060	7.55	496.00
Max	305.625	757.656	4,357,650

Correlation analysis was also conducted to evaluate relationships among variables in the multivariate dataset. The correlation matrix shows relatively weak linear relationships between temperature and pH (0.1513), temperature and TDS (−0.0011), and pH and TDS (−0.0272). Understanding inter-variable relationships is important in multivariate time-series forecasting because environmental parameters in aquaculture systems often interact dynamically over time and influence prediction performance[15].

Table 3. Correlation Matrix Analysis of Raw Water Quality Dataset

Variables of Raw Data	Temperature	pH	TDS
Temperature	1	0.15	-0.0010
pH	0.15	1	-0.0272
TDS	-0.0010	-0.0272	1

The results obtained from this exploratory correlation matrix analysis Table 3 were used as the basis for determining appropriate preprocessing strategies, including filtering, interpolation, normalization, and sliding window transformation, in order to improve dataset consistency and forecasting readiness.

3.3. Data Filtering

Data filtering was performed to ensure temporal consistency and reliability of the multivariate sensor dataset prior to sequence transformation. Since sensor measurements were recorded at 30-minute intervals, the dataset was first aligned to maintain uniform sampling resolution across all observation timestamps. Consistent temporal sampling is essential for multivariate time-series forecasting because irregular intervals may reduce prediction stability and affect sequence learning performance[16].

To support forecasting preparation, only the most recent 90 days of observations were selected from the monitoring records. This time window was chosen to represent the most relevant environmental dynamics of the aquaculture system while maintaining sufficient temporal continuity for supervised sequence construction. Similar fixed observation windows have been widely applied in environmental forecasting studies to improve model generalization and reduce the influence of outdated measurements. After the filtering process, the dataset contained 4,320 temporally consistent observations derived from three water quality parameters: temperature, pH, and total dissolved solids (TDS).

3.4. Missing Value Handling

To maintain dataset consistency, missing values in the filtered dataset were handled using interpolation techniques. Interpolation estimates unavailable observations based on neighboring temporal values, allowing the dataset to preserve sequential structure without removing valid timestamps.

This approach is commonly applied in environmental monitoring datasets to maintain continuity in multivariate time-series forecasting tasks and improve prediction stability. By reconstructing incomplete observations through interpolation, the resulting dataset becomes more suitable for sliding window transformation and multistep forecasting preparation. The interpolated value $x\ t$ x t between two known observations $x\ a$ x a and $x\ b$ x b is calculated as:

$$x_t = x_a + \frac{(x_b - x_a)}{(t_b - t_a)} \times (t - t_a) \quad (1)$$

Where $x\ t$ x t represents the interpolated observation at time t t, while $x\ a$ x a and $x\ b$ x b denote the nearest available observations before and after the missing value, respectively.

3.5. Outlier Detection

Outlier detection was performed to identify abnormal sensor readings that significantly deviated from normal environmental conditions. Extreme values in water quality monitoring datasets may originate from sensor instability, temporary environmental disturbances, or measurement noise, and may negatively affect forecasting performance if retained in the dataset.

$$Lower\ Bound = Q_1 - 1.5 \times IQR$$
$$Upper\ Bound = Q_3 + 1.5 \times IQR \quad (2)$$

The outliers were detected using the interquartile range (IQR) method based on percentile thresholds. The IQR approach identifies extreme observations located outside the lower and upper bounds defined by the first quartile (Q1) and third quartile (Q3), providing a robust mechanism for detecting anomalous values in environmental time-series datasets. This method is widely applied in sensor data preprocessing because it is resistant to distributional skewness and measurement noise. Removing extreme observations using the IQR-based filtering approach improves dataset stability and supports reliable multivariate sequence construction for forecasting preparation.

3.6. Data Normalization

Data normalization was applied to transform sensor observations into a consistent numerical scale prior to sequence construction. Since the dataset consists of multivariate environmental parameters with different measurement units and value ranges, normalization is required to prevent variables with larger magnitudes from dominating the learning process during multistep forecasting preparation.

In this study, Min–Max scaling was used to normalize temperature, pH, and total dissolved solids (TDS) observations into a uniform range between 0 and 1. This transformation preserves the relative distribution of each variable while improving numerical stability during multivariate sequence modeling. Min–Max normalization is widely applied in environmental time-series forecasting tasks because it supports stable convergence in neural-network-based prediction models and improves comparability across multiple sensor parameters.

The normalized value $x\ '$ x ′ is computed as:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

Where x x represents the original observation, $x_{\min}$ and $x_{\max}$ denote the minimum and maximum values of the variable, and x ′ is the normalized value within the interval [0,1]. This normalization process ensures that the dataset is suitable for multivariate sliding window transformation and subsequent forecasting sequence construction.

3.7. Sliding Window Transformation

Sliding window transformation was applied to convert sequential time-series observations into supervised learning sequences suitable for multistep forecasting tasks. This transformation organizes historical observations as input sequences and future observations as prediction targets, enabling forecasting models to learn temporal dependencies from structured multivariate sensor data[17].

In this research, each input sequence consisted of 144 timesteps representing historical observations collected at 30-minute intervals, corresponding to a three-day monitoring window. The prediction targets were also defined as the next 144 timesteps, allowing the dataset to support multi-step forecasting of temperature, pH, and total dissolved solids (TDS) simultaneously. This multivariate input–output structure enables forecasting models to capture inter-variable relationships and temporal dynamics across environmental parameters in aquaculture monitoring systems[18].

The transformation process converts the original time-series dataset into supervised sequence pairs expressed as:

$$X = \{x_{t-143}, x_{t-142}, y_{t-141}, \ldots, x_t\} \quad (4)$$

$$Y = \{y_{t+1}, y_{t+2}, y_{t+3}\}$$

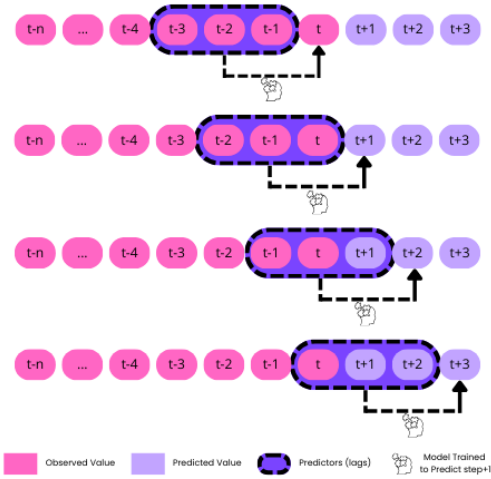

Figure 2. Framing Dataset in Sliding Window

In Fig. 2 can be seen Sliding Window usage, X represents the input sequence containing historical observations and Y represents the corresponding multi-step forecasting targets. This prediction process was repeated recursively until the complete forecasting horizon of 144 timesteps was achieved. The structured sequence construction improves dataset suitability for multivariate forecasting models and supports prediction of short-term water quality trends in aquaculture environments.

3.8. Data Splitting

After sliding window transformation, the structured multivariate sequences were divided into training, validation, and testing subsets to support forecasting model evaluation. The dataset was split using a ratio of 70% for training, 20% for validation, and 10% for testing to ensure balanced representation of temporal patterns across different stages of model development. From the total of 4,320 temporally consistent observations obtained after filtering, the dataset was proportionally allocated into 2,921 training sequences, 834 validation sequences, and 419 testing sequences, like Fig. 3.

This partitioning strategy ensures reliable evaluation of forecasting performance while preserving temporal dependencies within the multivariate sequence structure and the splitting process was performed sequentially to preserve temporal ordering between observations and prevent information leakage during forecasting evaluation.

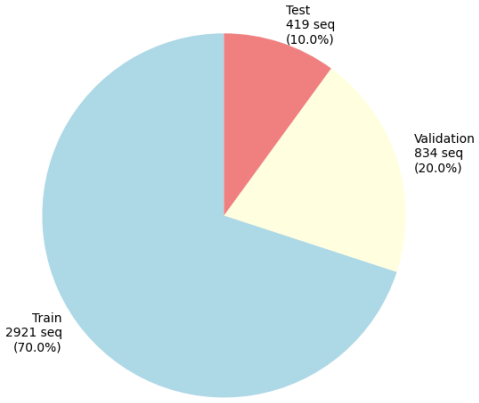

Figure 3. Split Distribution Visualization

4. RESULT AND DISCUSSIONS

4.1. Dataset After Preprocessing Steps

The descriptive statistics of the prepared dataset indicate that all variables were successfully transformed into a normalized range between 0 and 1, improving comparability across parameters with different measurement units. Compared with the raw dataset, which contained extreme minimum and maximum values outside realistic operational ranges, the normalized dataset shows more stable distributions across temperature, pH, and TDS observations. The reduced variability reflected by the standardized value ranges demonstrates that the preprocessing pipeline effectively minimized measurement noise and improved dataset suitability for multivariate time-series modeling. These improvements in Table 4 confirm that normalization plays an important role in preparing environmental sensor data for sequence-based forecasting tasks.

Table 4. Descriptive Statistics of Prepared Water Quality Dataset

Statistics	Temperature	pH	TDS
Count	4,320	4,320	4,320
Mean	0.5680	0.5178	0.4168
Std	0.1763	0.2823	0.2051
Min	0.0000	0.0000	0.0000
25%	0.4344	0.3523	0.3310
50% (Median)	0.5625	0.5114	0.4212
75%	0.7000	0.6856	0.5244
Max	1.0000	1.0000	0.9456

The correlation matrix obtained after preprocessing further illustrates the structural refinement of the multivariate dataset. While the raw dataset exhibited weak and inconsistent relationships between variables, the prepared dataset shows more interpretable inter-variable correlations, particularly between pH and TDS parameters. These results indicate that the preprocessing pipeline preserved meaningful environmental relationships while reducing the influence of abnormal sensor readings, as can be seen in Table 5. Maintaining stable inter-variable dependencies is essential for multivariate forecasting because prediction models rely on consistent temporal interactions among environmental parameters to generate reliable future estimates.

Table 5. Correlation Matrix Analysis of Prepared Water Quality Dataset

Variables of Raw Data	Temperature	pH	TDS
Temperature	1	0.0128	-0.1753
pH	0.0128	1	0.2109
TDS	-0.1753	0.2109	1

4.2. Sliding Window Sequences

The multivariate dataset was transformed into supervised learning sequences using a sliding window approach to support multi-step forecasting preparation. This transformation converts continuous temporal observations into structured input–output sequence pairs that enable prediction models to learn short-term environmental dynamics from historical sensor measurements. In this study, each input sequence consisted of 144 timesteps representing historical observations collected at 30-minute intervals, corresponding to a three-day monitoring window.

$$X = \begin{bmatrix} Temp_{t-143} & pH_{t-143} & TDS_{t-143} \\ \vdots & \ddots & \vdots \\ Temp_t & pH_t & TDS_t \end{bmatrix}$$

$$Y = \begin{bmatrix} Temp_{t+1} & pH_{t+1} & TDS_{t+1} \\ Temp_{t+2} & pH_{t+2} & TDS_{t+2} \\ Temp_{t+3} & pH_{t+3} & TDS_{t+3} \end{bmatrix} \quad (5)$$

The prediction targets were defined as the subsequent 144 timesteps, allowing simultaneous multi-step forecasting of temperature, pH, and total dissolved solids (TDS). This sequence construction enables prediction models to capture both temporal dependencies and inter-variable relationships across multiple environmental parameters within the monitoring system. Formally, the sliding window transformation converts the normalized dataset into multivariate input–output matrices expressed as:

Where N represents the number of generated sequence pairs, 144 denotes the number of timesteps in input sequence window while output has 3 timesteps prepared for recursive prediction, and 3 corresponds to the number of monitored parameters (temperature, pH, and TDS).

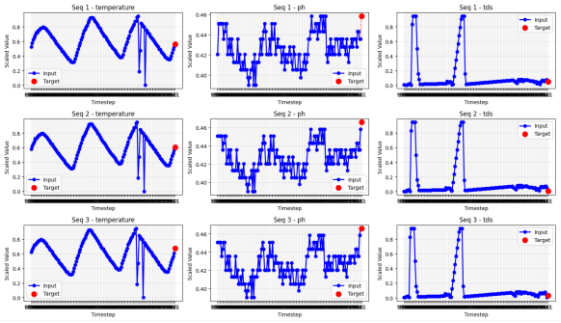

Figure 4. The First Three Sequences Input in Prepared Dataset in 144 Data Points

Fig. 4 is the examples of the constructed input sequences illustrate how historical observations from the previous three days are used to predict environmental conditions over the next three days. This structured representation improves dataset suitability for multivariate sequence-based forecasting and supports prediction of short-term variations in water quality conditions.

4.3. Forecasting Validation Evaluation

To evaluate the forecasting readiness of the prepared dataset, a multivariate Long Short-Term Memory (LSTM) network was applied as a validation model for sequence-based prediction of water quality parameters. LSTM architectures are widely used in environmental time-series forecasting because they are capable of capturing nonlinear temporal dependencies and long-term sequential relationships among multiple monitoring variables, making them suitable for water quality prediction tasks[19]. The validation model consisted of two stacked LSTM layers with 256 and 128 neurons, respectively, followed by dropout layers with a dropout rate of 0.2 to reduce overfitting during sequence learning. A dense layer combined with a reshape operation was applied at the output stage to support multivariate multi-step

prediction across temperature, pH, and total dissolved solids (TDS) parameters[20]. This architecture enables simultaneous forecasting of multiple environmental variables over the defined prediction horizon.

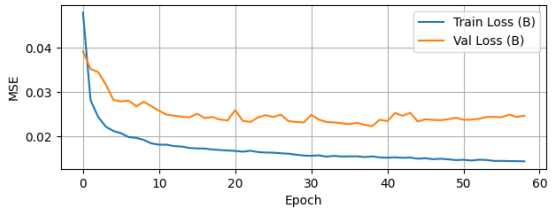

Figure 5. Training Validation Loss Visualization

Fig. 5 is the convergence behavior of the training process was evaluated using training and validation loss curves observed during model optimization. The results indicate a consistent decrease in both training and validation loss values across epochs, demonstrating stable learning performance without significant divergence between the two curves. This pattern suggests that the normalized dataset preserved sufficient temporal structure and inter-variable consistency to support effective sequence learning while minimizing the risk of overfitting. The observed convergence behavior confirms that the preprocessing pipeline successfully produced structured multivariate sequences suitable for recursive forecasting tasks.

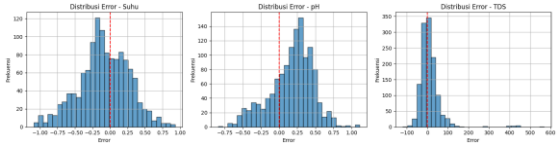

Figure 6. Error Distribution Every Parameters

The error distribution analysis further confirms the stability of the forecasting results obtained using the prepared dataset. As illustrated in Fig. 6, the prediction errors for temperature and pH are approximately centered around zero, indicating that the model predictions do not exhibit significant systematic bias. Although the error distribution for TDS shows a wider spread compared to the other parameters, the majority of prediction errors remain concentrated near zero, suggesting that the recursive forecasting process preserves the overall temporal structure of the dataset. These results support the effectiveness of the proposed preprocessing pipeline in producing structured multivariate sequences suitable for short-term water quality prediction.

Table 6. Evaluation Metrics Result By Multivariate LSTM Model

Parameters	MAE	RMSE	R^2
Temperature	0.2936	0.3690	0.9405
pH	0.2915	0.3432	0.2571
TDS	30.9269	56.5534	0.9058

The evaluation results further demonstrate the suitability of the prepared dataset for multivariate forecasting applications. As shown in Table 6, the model achieved strong prediction performance for temperature and TDS parameters with coefficient of determination values of 0.9405 and 0.9058, respectively, indicating that the constructed sequences effectively preserved temporal dynamics of environmental observations. Although the pH parameter showed relatively lower predictive accuracy with an R^2 value of 0.2571, this behavior is commonly observed in water quality monitoring datasets due to higher variability and sensitivity of pH measurements.

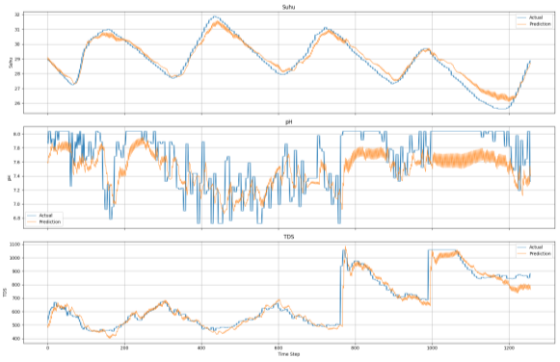

Figure 7. Data Test Prediction with Ground Truth Data in 3 Days

Furthermore, the correlation structure observed after preprocessing, as illustrated in Fig. 7, indicates that the relationships among temperature, pH, and TDS parameters remain sufficiently preserved for multivariate forecasting purposes. Although the correlation values between variables are relatively moderate, they reflect meaningful interactions commonly found in environmental monitoring datasets where water quality parameters influence each other dynamically over time. The preservation of these inter-variable dependencies confirms that the preprocessing

pipeline maintained essential temporal relationships required for sequence-based prediction and supports the suitability of the prepared dataset for multivariate LSTM forecasting applications.

5. CONCLUSION

a. This study proposed a structured preprocessing pipeline for multivariate water quality sensor data to improve dataset readiness for time-series forecasting in aquaculture monitoring environments. The pipeline includes EDA, temporal filtering, interpolation for missing values, outlier removal (IQR), Min–Max normalization, sliding window transformation, and dataset splitting, resulting in a consistent dataset of 4,320 observations.
b. The preprocessing process improves data quality by reducing noise and preserving meaningful inter-variable relationships among temperature, pH, and TDS parameters. Sliding window transformation enables the construction of multivariate input–output sequences with a 144-timestep forecasting horizon for effective short-term prediction. Validation using an LSTM model demonstrates stable forecasting performance, confirming that the proposed framework produces reliable and forecasting-ready datasets

GRADITUTE WORDS

The author would like to express sincere gratitude to the Living Lab Smart Aquaculture initiative organized by the Faculty of Engineering and Vocational for providing the research platform and facilities, as well as to the supervising lecturer for the continuous guidance, support, and valuable insights throughout the research and writing process. Special thanks are addressed to colleagues and team members who have contributed, supported, and collaborated in developing this project, making it grow into a meaningful and impactful work ready to be implemented in real-world applications. Finally, the author would like to thank family and relatives for their unwavering support, encouragement, and understanding, which made the completion of this research possible, as well as for the emotional support that has been invaluable throughout this journey.

REFERENCES

[1] B. G. K. Yudistira, C. Hapsari, G. D. W. Adnyana, W. Nath, and I. P. R. M. Putra, “Smart Fisheries: Real-Time Water Quality Management and Automated Feeding System Design for Tilapia Farming using ESP32 Micro Controller,” Journal of Artificial Intelligence and Software Engineering (J-AISE), vol. 5, no. 2, pp. 827–832, Jun. 2025, doi: 10.30811/jaise.v5i2.7288.

[2] Gandh, R. D., Harigovindan, V. P., Rasheed Abdul Haq, K. P., & Bhide, A. (2024). Attention-driven LSTM and GRU deep learning techniques for precise water quality prediction in smart aquaculture. Aquaculture International. https://doi.org/10.1007/s10499-024-01574-5

[3] P. Liu, J. Wang, A. K. Sangaiah, Y. Xie, and X. Yin, “Analysis and Prediction of Water Quality Using LSTM Deep Neural Networks in IoT Environment,” Sustainability, vol. 11, no. 2058, pp. 2–14, Apr. 2019, doi: 10.3390/su11072058.

[4] S. C. M. Sundararajan et al., “IoT-based prediction model for aquaponic fish pond water quality using multiscale feature fusion with convolutional autoencoder and GRU networks,” Sci. Rep., vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-024-84943-7.

[5] R. Krishnamurthi, A. Kumar, D. Gopinathan, A. Nayyar, and B. Qureshi, “An overview of iot sensor data processing, fusion, and analysis techniques,” Nov. 01, 2020, MDPI AG. doi: 10.3390/s20216076.

[6] K. P. Rasheed Abdul Haq and V. P. Harigovindan, “Water Quality Prediction for Smart Aquaculture Using Hybrid Deep Learning Models,” IEEE Access, vol. 10, pp. 60078–60098, 2022, doi: 10.1109/ACCESS.2022.3180482.

[7] R. Baena-Navarro, Y. Carriazo-Regino, F. Torres-Hoyos, and J. Pinedo-López, “Intelligent Prediction and Continuous Monitoring of Water Quality in Aquaculture: Integration of Machine Learning and Internet of Things for Sustainable Management,” Water (Switzerland), vol. 17, no. 82, pp. 1–25, Jan. 2025, doi: 10.3390/w17010082.

[8] E. Antony, N. S. Sreekanth, R. K. Sunil Kumar, and T. Nishanth, “Data preprocessing techniques for handling time series data for environmental science studies,” International Journal of Engineering Trends and Technology, vol. 69, no. 5, pp. 196–207, May 2021, doi: 10.14445/22315381/IJETT-V69I5P227.

[9] E. Eze, S. Kirby, J. Attridge, and T. Ajmal, "Aquaculture 4.0: Hybrid Neural Network Multivariate Water Quality Parameters Forecasting Model," Res. Sq., pp. 1–21, Mar. 2023, doi: 10.21203/rs.3.rs-2711537/v1.

[10] P. Martín-Calzada, P. Martín Sánchez, F. J. Rodríguez-Sánchez, C. Santos-Pérez, and J. Ballesteros, "Optimized Sensor Data Preprocessing Using Parameter-Transfer Learning for Wind Turbine Power Curve Modeling," Sensors, vol. 25, no. 17, Sep. 2025, doi: 10.3390/s25175329.

[11] J. Wang, W. Jiang, Z. Li, and Y. Lu, "A new multi-scale sliding window lstm framework (Mssw-lstm): A case study for gnss time-series prediction," Remote Sens. (Basel)., vol. 13, no. 16, Aug. 2021, doi: 10.3390/rs13163328.

[12] N. Okafor and D. Delaney, "MISSING DATA IMPUTATION ON IOT SENSOR NETWORKS: IMPLICATIONS FOR ON-SITE SENSOR CALIBRATION," Jan. 26, 2021, techrxiv by IEEE. doi: 10.36227/techrxiv.13633529.v1.

[13] R. Chandra, S. Goyal, and R. Gupta, "Evaluation of Deep Learning Models for Multi-Step Ahead Time Series Prediction," IEEE Access, vol. 9, pp. 83105–83123, 2021, doi: 10.1109/ACCESS.2021.3085085.

[14] K. W. Palihakkara and M. N. Jayakody, "A Comparative Study on Deep Learning Models for Time-Series Forecasting," Jan. 26, 2026, preprints.org. doi: 10.20944/preprints202601.1962.v1.

[15] X. Ji et al., "Long-term multivariate water quality forecasting for sustainable aquaculture management," Water Res. X, vol. 29, Dec. 2025, doi: 10.1016/j.wroa.2025.100402.

[16] Z. Li, S. Qi, Y. Li, and Z. Xu, "Revisiting Long-term Time Series Forecasting: An Investigation on Linear Mapping," ArXiv, May 2023, [Online]. Available: http://arxiv.org/abs/2305.10721

[17] J. Brownlee Disclaimer, Deep Learning for Time Series Forecasting Predict the Future with MLPs, CNNs and LSTMs in Python. Machine Learning Mastery, 2018.

[18] A. Jaffar, N. M. Thamrin, M. S. A. M. Ali, M. F. Misnan, and A. I. M. Yassin, "WATER QUALITY PREDICTION USING LSTM-RNN: A REVIEW," J. Sustain. Sci. Manag., vol. 17, no. 7, pp. 204–225, Jul. 2022, doi: 10.46754/jssm.2022.07.015.

[19] I. P. R. Mahaputra, B. G. K. Yudistira, J. Shiddiq, G. D. W. Adnyana, and P. Z. E. S. Nugraha, "Short-term time-series forecasting of hydroponics water quality parameters using XGBoost based IoT sensor data," JITET (Jurnal Informatika dan Teknik Elektro Terapan), vol. 14, no. 2, 2026, doi: 10.23960/jitet.v14i2.9458.

[20] Y. Yang, P. Zhang, and Y. Wang, "An attention-based parallel model with sliding window decomposition algorithm for water quality prediction," Journal of Water Process Engineering, vol. 78, p. 108751, 2025, doi: 10.1016/j.jwpe.2025.108751.